

Feature Extraction from Electroencephalograms for Bayesian Assessment of Newborn Brain Maturity

Livia Jakaite, Vitaly Schetinin, *Joachim Schult
University of Bedfordshire, UK, *Universität Hamburg, Germany
livija.jakaite@beds.ac.uk

Abstract

We explored the feature extraction techniques for Bayesian assessment of EEG maturity of newborns in the context that the continuity of EEG is the most important feature for assessment of the brain development. The continuity is associated with EEG “stationarity” which we propose to evaluate with adaptive segmentation of EEG into pseudo-stationary intervals. The histograms of these intervals are then used as new features for the assessment of EEG maturity. In our experiments, we used Bayesian model averaging over decision trees to differentiate two age groups, each included 110 EEG recordings. The use of the proposed EEG features has shown, on average, a 6% increase in the accuracy of age differentiation.

1. Introduction

Abnormal brain development in newborns is a life-threatening factor, and early diagnosis of such development is important for saving health and lives of newborns [1], [2]. Clinical assessment of the brain development is mainly carried out by analyzing the maturity-related features in sleep electroencephalograms (EEGs). Assessments are typically made in terms of a newborn’s physiological age counted by the number of weeks post-conception. Based on the clinical evidences, the post-conceptual and EEG-estimated ages are typically matched if a newborn is healthy, and when the age estimate is mismatched, the newborn’s brain development is unlikely normal [2]. However, the maturity-related features are difficult to recognize and assess as they widely vary between patients. Such variations make the EEG-based reference guides inconsistent for clinical use. The computer-aided analysis of the maturity-related patterns in sleep EEGs can improve the care of newborns at high-risk [3], [4].

Typically, computer-aided analysis of sleep EEG is carried out with the spectral powers computed in the standard frequency bands [3]. The use of such EEG features has been shown promising for assessment of EEG maturity [1], [15]. To cope with the non-station-

arity of EEG the spectral analysis has been carried out within short, up to 10s, intervals [3]. Within this approach, no attempts have been made to measure or estimate the level of “stationarity” of EEG which is an important maturity-related feature recognizable by experts as a pattern of the background EEG continuity [3].

Patterns of EEG continuity (or *vice versa* discontinuity) are difficult to recognize as they widely vary over time and between patients. By definition, an EEG pattern is discontinuous if the intervals with the normal voltage range are interchanged with periods of low voltage. At 28 to 30 weeks post-conception, EEG intervals comprise high-amplitude bursts of the mixed-frequency waves interrupted by long low-voltage periods, named *inter-burst* intervals [3]. During the brain development, longer periods of uninterrupted EEG activity appear, and the continuity is progressively increased [3], [5], [6].

For automated estimation of the discontinuity in newborn EEG, threshold segmentation has been proposed in [7]. However, this technique requires to set a proper threshold for each newborn EEG recording. Adaptive segmentation, proposed in [8], [9], [12] to mitigate this problem, aimed to split EEG into *pseudo-stationary* intervals in which the EEG amplitudes and frequencies vary within an acceptable small range. These intervals are then used for extracting EEG features.

Based on the above clinical observations, in this paper we propose a new method for extracting features from newborn EEG. We expect that the proposed method is capable of distinguishing longer pseudo-stationary segments within the continuous patterns against relatively shorter segments within the discontinuous patterns. There have been found clinical evidences that the longer pseudo-stationary segments are detected in EEG of more mature newborns. Consequently, we can assume that the rate of pseudo-stationary intervals segmented in an EEG is an important feature of the age-related pattern.

Our assumption are tested within the Bayesian classification of two age groups represented by the stand-

ard spectral power bands extended with the newly proposed features. We show first that the proposed features are highly correlated with the post-conceptual age, and second that the use of these features provides more accurate classification of post-conceptual age.

The classification is implemented within the methodology of Bayesian Model Averaging (BMA) over Decision Trees (DTs) that enables users to evaluate the uncertainty of classification [10]. The use of DTs enables to evaluate the contribution of features to the classification.

The rest of the paper is structured as follows. Section 2 briefly introduces the methodology of BMA over DTs, and Section 3 describes a technique proposed for EEG segmentation into pseudo-stationary intervals. Section 4 describes the experiments with the new EEG features and the Bayesian classification of newborn ages. Finally Section 5 concludes the paper.

2. Bayesian Model Averaging over Decision Trees

In theory, Bayesian averaging over multiple models provides the most accurate estimates of class posterior probabilities, which then can be used to evaluate the uncertainty in the model outcomes, and particularly to count the the false positive and false negative rates.

The use of BMA for classification has been made feasible by using Markov Chain Monte Carlo (MCMC) stochastic approximation (see e.g. [10]). The MCMC has been extended to approximate the posterior distribution from classification models of a variable dimensionality by using Reversible Jump (RJ) described in [11]. The RJ MCMC provides the conditions under which the estimate of class posterior becomes most accurate.

The use of DT models enables to evaluate the contribution of features to the classification, and therefore the BMA over the DTs will benefit us with the information about feature importance. The example of the implementation of BMA over DTs has been described in our previous research [13].

To let a DT model grow, the RJ MCMC makes “birth”, “death”, and change moves. The first two moves change the number of DT splitting nodes (and the model dimensionality), and the other moves change only the DT parameters. All modifications are proposed within the given prior distributions as follows.

The birth move randomly splits the data points falling in one of the terminal nodes by a new splitting node with the variable and rule drawn from a given probability function.

The death move randomly picks a splitting node with two terminal nodes and assigns it to be one terminal with the united data points.

The change-split move randomly picks a splitting node and then assigns it a new splitting variable and rule drawn from a given probability function.

The change-rule move randomly picks a splitting node and then assigns it a new rule drawn from a given prior.

A modification proposed for DT parameters is evaluated in terms of likelihood of the model. The modification is accepted or rejected accordingly to the Bayes’ rule. If the proposed changes improve the likelihood of the current model, they are always accepted. If else, the proposed model parameters are accepted with a smaller probability.

During MCMC simulation, a DT grows and its likelihood tends to grow. This phase is called *burn-in* and it should be preset sufficiently long in order to reach a stable DT. When a DT has been grown, its likelihood becomes stable. During this phase named *post burn-in*, samples from the posterior distribution are collected to be averaged.

3. Adaptive segmentation of EEG

This section briefly describes the underlying concepts of adaptive segmentation of EEG which were used for the proposed technique. The details discussed below are needed to implement the segmentation techniques for our comparative experiments.

3.1. Conventional techniques

As discussed in the Introduction, the aim of adaptive segmentation is to automatically label boundaries of pseudo-stationary intervals in EEG. The segment boundaries are typically found by evaluating the dissimilarity of EEG in one or more successive intervals. Typically, the dissimilarity is evaluated between two intervals called reference and test windows. If its value is small, samples of the reference and test windows are considered taken from a similar stationary process. On the contrary, if the dissimilarity exceeds a given threshold, the samples are assumed taken from the different processes, and a segment boundary is assigned between the reference and test windows. The reference and test windows are typically made adjoined and sliding along the EEG. Alternatively, the reference window can be fixed at the beginning of the segment while the test window is moved until the dissimilarity exceeds a given threshold [8].

The dissimilarity is often evaluated as the difference between coefficients of auto-regression models fitted to the reference and test windows. This technique requires to find the proper threshold as well as the proper number of the coefficients.

An adaptive EEG segmentation technique proposed in [12] aimed to estimate the dissimilarity between the

reference and test windows by using the Nonlinear Energy Operator (NLEO) enabling to estimate “weighted energy”. Within this technique, a threshold is automatically found in a sliding window of a predefined size. However, the proper length of the threshold adaptation window has to be found from experiments.

Both techniques attempt to assess the dissimilarity of EEG intervals in terms of wave frequencies and energy of the signal. Despite that the thresholds are adapted to the EEG intervals, these techniques tend to fail segmentation of intervals with age-related events which are characterized by abrupt changes in the frequencies and amplitudes of the signal.

3.2. The proposed technique

Alternatively to the above techniques, we propose to segment EEG by using estimates of the *spectral* dissimilarity of EEG intervals. In our method, the dissimilarity is assessed over spectral power bands computed in the reference and test windows. The spectral estimates are then compared within a *two-sample hypothesis test*. Thus, the main idea behind our method is to employ Spectral Power Statistics (SPS) for making decisions on the dissimilarity of EEG intervals.

Typically, the spectral powers are calculated for the following six frequency bands: *Subdelta*, *Delta*, *Theta*, *Alpha*, *Beta*, and *Beta2*. Within our method, this set is slightly extended in order to met the requirements of a statistical test. Particularly, the number of bands is increased to nine to satisfy the requirement of the standard tests based on the Kolmogorov-Smirnov statistics over cumulative distributions.

Based on our observations, we assumed that each of *Theta*, *Alpha* and *Beta* bands could be split into sub-bands. In our experiments, these additional sub-bands have been shown increasing the accuracy of EEG assessment.

In the comparison with the conventional segmentation techniques, the proposed technique needs to define a smaller number of parameters. These parameters mainly include the following: (1) a structure of spectral bands based on their conventional structure, (2) durations of the reference and test windows, (3) a threshold level of rejecting the alternative hypothesis, and (4) a structure of distribution of the tested samples.

These parameters are slightly dependent on each other, so that they can be optimized from experiments. Under the above conditions, the proposed technique is expected to segment sleep EEG into pseudo-stationary intervals more accurately than the conventional techniques within a wide range of variations between patients.

In the next section, we describe experiments in which the proposed technique is compared with the conventional segmentation techniques, and the new

features extracted from segmented EEG intervals are used for assessment of EEG maturity.

4. Experiments and results

In our experiments, we used sleep EEGs recorded from healthy newborns via the standard C3T3 and C4T4 electrodes with sampling rate of 100 Hz. First, we evaluated the stationarity of EEGs over 11 age groups, and second, we ran experiments with the Bayesian assessment of EEG maturity.

4.1. The evaluation of EEG stationarity

The experiments have been run on 220 EEGs recorded from newborns of 11 age groups between 34 and 44 weeks post-conception, with 20 recordings in each group. The stationarity of an EEG recording has been evaluated by a histogram of pseudo-stationary intervals detected by a segmentation technique. The total sum over the histogram counts has been used to estimate the rate of the pseudo-stationary intervals segmented in an EEG recording. The larger the rate, the higher is the level of EEG non-stationarity.

In our experiments, we used a number of segmentation techniques and found that the rates of non-stationarity are correlated with the age post-conception differently. Table 1 shows the relationships between the rates and ages, represented by the Spearman rank correlation coefficient ρ and significance p . The coefficients were computed for the conventional amplitude statistics (AS) and the proposed SPS-based technique described in Section 3.2.

The first segmentation technique employs the NLEO applied to the reference and test windows, each of 1-s duration. The duration of the threshold adaptation window sliding over the EEG was two seconds. The correlation between the rate of non-stationarity and age was insignificant ($\rho = -0.07$, and confidence probability $p = 0.28$) as the numbers of pseudo-stationary intervals were similar for all 11 age groups. This likely happens because this technique assigns the maximal energy in the threshold adaptation window to be a threshold, so that pseudo-stationary intervals are most likely broken within the 2-s threshold adaptation window.

Table 1. Correlation coefficients between the post-conceptual age and rate of intervals

Technique	AS	SPS
NLEO	-0.07	-
AR	-0.57	-
X^2	0.24	-0.67
t -test	0.28	-0.75
K-S	0.17	-0.75
A-D	-0.22	-0.81

The second EEG segmentation technique employs the auto-regression (AR) models computed in the reference and test windows of 2-s duration. This technique starts a new interval when the cross-validation errors of the AR models exceeded a given threshold. The correlation between the rate of intervals and ages was much higher ($\rho = -0.57$, and $p < 0.0001$).

We also compared the standard two-sample X^2 , t -test, Kolmogorov-Smirnov (K-S), and Anderson-Darling (A-D) tests. The reference and test windows were adjoined, and their durations were set to two seconds. We applied these tests to the both statistics, the conventional AS and the proposed SPS, to compare their informativeness in terms of correlation between rates of EEG non-stationarity and newborn ages.

The spectral powers required by the SPS were computed with the discrete Fourier transform and then summed within the standard frequency bands, whose number has been increased from six to nine as described in Section 3.2.

Both t -test and K-S test have shown the similar correlation, $\rho = -0.75$, although t -test is typically applied to samples of a normal distribution, and K-S test employs a cumulative distribution making this test more robust to variations of distribution shape as well as to variations of the power of EEG.

A modification of K-S test known as A-D test tends to assign larger weight to the tails of distributions. In our experiments, this test has shown providing a better correlation ($\rho = -0.81$, and $p < 0.0001$).

The above four test techniques have been applied to the AS over EEG with the same window durations. The highest correlation has been achieved with the t -test ($\rho = 0.28$, and $p < 0.0001$). We can therefore conclude that the segmentation techniques based on the AS presentation are less informative than those based on the proposed SPS-based segmentation of EEG.

In our experiments we used sleep EEGs which are typically contaminated by muscle and electrode artifacts, which can affect the informativeness of features extracted from the EEG. For this reason, we were interested in investigating how artifact removal can improve the results of the K-S, A-D and t -tests employing the proposed SPS-based segmentation technique.

For these experiments, we used the two techniques of removal of amplitude artifacts from sleep EEG. The first technique (AR1) simply removes EEG samples with amplitudes exceeding a threshold assigned as the sum of the mean and standard deviation counted in a 10-min sliding window. The second technique (AR2) employs the NLEO to remove EEG samples of abnormal energy detected in a 3-min sliding window. The rates of artifacts removed by these techniques were similar.

Table 2 shows the correlation coefficients after the

artifact removal, and we can observe that the first technique (AR1) has been shown increasing the correlation for the K-S and A-D tests by 3%. The second technique (AR2) has not been shown improving the correlation.

Table 2. Correlation between the post-conceptual age and rate of segments detected by the K-S, A-D and t -test after removal of EEG artifacts

	Artifacts, %	K-S	A-D	t -test
AR1	0.28 ± 0.08	-0.84	-0.84	-0.76
AR2	0.29 ± 0.20	-0.71	-0.79	-0.7

Fig. 1 shows the rates of EEG non-stationarity estimated by the K-S test after the artifacts removal by the technique AR1 for newborns at age between 34 and 44 weeks post-conception.

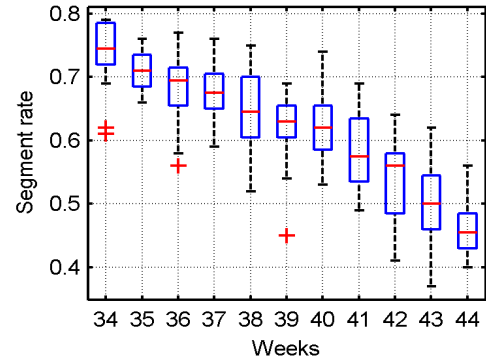


Figure 1. Correlation between the post-conceptual age and the segment rates after removal of EEG artifacts

In the next section, we describe experiments with the BMA over DTs using the new EEG features for newborn age classification.

4.2. Bayesian classification of EEG maturity

In our experiments, we used the BMA over DTs using features extracted from EEG for classification of EEG maturity of newborns at ages 36 and 41 (36/41), and 37 and 39 (37/39) weeks post-conception. Each age group was represented by 110 sleep EEG recordings. As the EEG maturity of preterm (36 weeks) and full-term (41 weeks) newborns is different, the accuracy of classification of these EEGs is expected high. In contrast, we expect that EEG maturity patterns of newborns at ages of 37 and 39 weeks are more close, and the classification accuracy is expected lower.

Table 3. Performances of age classification with the different sets of features

	Set 1		Set 2		Set 3	
Weeks	P , %	E	P , %	E	P , %	E
36/41	83.6±8.7	24.2±6.6	87.3±12.3	16.9±6.0	90.0±9.4	15.0±3.5
37/39	64.5±6.1	38.4±2.7	73.2±12.6	35.9±3.0	74.5±10.8	34.0±4.9

The EEGs have been cleaned from the amplitude artifacts and then segmented into pseudo-stationary intervals by using the proposed technique. In the experiments, we used three different sets of EEG features.

The Set 1 has comprised the conventional 36 spectral features which are the absolute and relative powers computed in the standard six frequency bands for the C3 and C4 electrodes and their sum.

The Sets 2 and 3 have comprised the features included in Set 1 and the new features representing the non-stationarity of an EEG recording estimated with the proposed SPS-based techniques which employ the K-S and A-D tests, respectively. The new features includes the segment rate and 10-bin histograms of the pseudo-stationary intervals ranging from 2s to 20s.

Table 3 shows the average performances P and entropies E of DT ensembles along with 2σ intervals obtained for the Bayesian classification of EEG of newborns at ages 36/41 and 37/39. The performances were evaluated within the 5-fold cross-validation.

The BMA over DTs has been implemented as described in Section 2 and run with the following settings. The length of the burn-in and post burn-in phases was set to 100,000 and 10,000, respectively. The minimal number of data samples allowed to be in DT nodes was 2. The proposal variance was given 1.0, and the probabilities of birth, death, change-split, and change-rule moves were 0.15, 0.15, 0.1 and 0.6, respectively.

Under these setting, the acceptance rate was around 0.4 in both phases, and the DT became stationary after growing to six nodes. Table 3 shows the performances of the BMA over DTs employing the Sets 1, 2, and 3, respectively. We can observe that for classification of ages at 36 and 41 weeks the features of Set 3 improve the performance, on average, by 6.4%, and Set 2 only by 3.7%. For classification of ages at 37 and 39 weeks, the gains are 10.0%, and 8.7%, respectively.

Besides, the new features included in Set 3 decrease the uncertainty of DT ensembles in term of entropy. The mean entropy of the DT ensembles has been reduced from 24.2 to 15.0 for classification of newborns at 36 and 41 weeks, and from 6.1 to 4.9 for classification of newborns at 37 and 39 weeks.

5. Conclusions and Discussion

We explored the feature extraction techniques for the Bayesian assessment of EEG maturity in full and preterm newborns in the context that the continuity of the EEG is the most influential feature in assessment of the brain development. We introduced EEG “stationarity” which is assessed by adaptive segmentation of the EEG into pseudo-stationary intervals. Our assumption is based on the clinical observations that the continuous EEG patterns with mixed frequencies and low variation in amplitude become longer, while discontinuous patterns with higher amplitude variation and slow activity in bursts become shorter during the brain development. We expected that longer pseudo-stationary segments would be correlated with the continuous patterns, and shorter segments with the discontinuous patterns. Thus we hypothesized that the rate of the pseudo-stationary intervals is highly correlated with the EEG maturity of newborns.

We proposed the new Spectral Power Statistics (SPS) technique testing the hypothesis of similarity between the distributions of spectral powers in the adjoined EEG intervals. We explored the robustness of this technique to EEG artifacts and found that the rate of pseudo-stationary intervals is highly correlated with EEG maturity ($\rho = -0.84$, and $p < 0.0001$).

The performances of the proposed SPS-based segmentation techniques have been explored within the conventional X^2 , t -test, Anderson-Darling (A-D) and Kolmogorov-Smirnov (K-S) tests. We found that the A-D test and K-S tests provide a better performance in terms of the correlation coefficient value.

The histograms of the pseudo-stationary intervals were used to extend the conventional set of spectral features for the Bayesian assessment of EEG maturity of newborns at ages 36 and 41 weeks as well as 37 and 39 weeks. The use of the new features has been shown increasing the accuracy of age classification by 6.3% and 10.0%, respectively, for these age groups. Overall, the classification accuracies have been increased to 90.9% and 75.4%, respectively.

Acknowledgments. The research is funded by the Leverhulme Trust, UK. The authors are grateful to the University of Jena, Germany, for granting the clinical EEG data for the research, and to the anonymous reviewers for their useful comments on the paper.

6. References

- [1] K. Holthausen et al. "Clinical relevance of age-dependent EEG signatures in the detection of neonates at high risk for apnea," *Neuroscience Letters Volume*, vol. 268, iss. 3, pp. 123–126, 199.
- [2] M. S. Scher, "Neurophysiological assessment of brain function and maturation: A measure of brain adaptation in high risk infants," *Pediatric Neurology*, vol. 16, iss. 3, 1997.
- [3] R. Cooper, C. Binnie, and J. C. Schaw, "Clinical neurophysiology: EEG, paediatric neurophysiology, special techniques and applications," , C. Binnie, R. Cooper, F. Mauguire, J. W. Osselton, P. F. Prior, and B. M. Tedman, Eds. Elsevier Science, 2003.
- [4] M. S. Scher, J. Turnbull, K. Loparo, and M. W. Johnson, "Automated state analyses: proposed applications to neonatal neurointensive care," *Journal of Clinical Neurophysiology*, vol. 22, iss. 4, pp. 256–270, 2005
- [5] M. Olischar et al. "Reference values for amplitude-integrated electroencephalographic activity in preterm infants younger than 30 weeks" gestational age," *Pediatrics*, vol. 113, iss. 1, pp. 61–66, 2004.
- [6] G. B. Boylan, D. M. Murray, and J. M. Rennie, "The normal EEG and aEEG," in *Neonatal cerebral investigation*, J. M. Rennie, C. F. Hagmann, and N. J. Robertson, Eds. UK: Cambridge University Press, 2008, pp. 83–91.
- [7] C. R. West et al. "Cot-side electroencephalography for outcome prediction in preterm infants: observational study," *Archives of disease in childhood. Fetal and neonatal edition*. vol. 96, iss. 2, p. F108–13, 2011.
- [8] R. Aufrichtig, S. B. Pedersen, and P. Jennum, "Adaptive segmentation of EEG signals," in *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1991, pp. 453–454.
- [9] L. Wong, "Quantitative Continuity Feature for Preterm Neonatal EEG Signal Analysis," PhD Thesis, The University of Auckland, 2008.
- [10] H. Chipman, E. George, and R. McCulloch, "Bayesian CART model search," *Journal of American Statistics*, vol. 93, pp. 935–960, 1998.
- [11] P. J. Green. "Reversible Jump Markov Chain Monte Carlo and Bayesian Model Determination", *Biometrika*, 82, pp 711-732, 1995.
- [12] R. Agarwal, J. Gotman, D. Flanagan, and B. Rosenblatt, "Automatic EEG analysis during long-term monitoring in the ICU," *Electroencephalography and Clinical Neurophysiology*, vol. 107, iss. 1, pp. 44–58, 1998.
- [13] L. Jakaite et al. "Bayesian decision trees for EEG assessment of newborn brain maturity," in *Proc. UKCI 2010, the 10th Annual Workshop on Computational Intelligence*, 2010.
- [14] D. M. Mason and J. H. Schuenemeyer, "A Modified Kolmogorov-Smirnov Test Sensitive to Tail Alternatives," *Annals of Statistics*, vol. 11, iss. 3, pp. 933–956, 1983.
- [15] M. S. Scher, D. A. Steppe, and D. L. Banks, "Prediction of lower developmental performances of healthy neonates by neonatal EEG-sleep measures," *Pediatric Neurology*, vol. 14, iss. 2, pp. 137–144, 1996.